CHARACTERISTICS OF GRADING MODELS

David Edward Polzien

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

Characteristics of Grading Models

by

David Edward Polzien

Thesis Advisor: D. R. Barr

December 1972

Approved for public release; distribution unlimited.

Characteristics of Grading Models

by

David Edward Polzien
Ensign, United States Navy
B.S. United States Naval Academy, 1971

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 1972

ABSTRACT

Many questions have arisen concerning grading systems. From a review of the current literature, it is apparent that very little mathematical analysis has been completed in this area. Since the grading process as most often used is a mathematical process, it seems desirable to examine grading systems from the mathematical point of view.

In this paper the familiar five-letter grading system is modeled mathematically. Variance of the grade point average is used as the measure of effectiveness. The grading system and its parameters are defined and the effect of variation in the parameters of the student performance distribution is observed. The effect of having multiple graders in a grading system is also modeled as is the effect of changing the number of grading categories in a grading system.

Grade data was obtained from the records of a number of students who attended the Naval Postgraduate School and an analysis of this data is presented.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

A. BACKGROUND

During the past few years a number of colleges and universities have modified their grading systems. This experimentation has come about because of a feeling that earlier grading systems are inadequate. Many aspects of grading systems have come under attack as not doing what they should do.

At present the grading system in predominant use is the five-letter system with a numerical value assigned to each letter which allows combination of grades into a grade point average. The grade point average is often used for determination of class standing or ranking the students in order. Alternative evaluation systems have been proposed to replace this basic system and its variations. One of the most radical systems proposed is that of independent study with no determination of grade being made, not even pass or fail. This system is not likely to replace the current systems in the near future, yet consideration of such a system points out that even the basic concepts concerning student evaluation are being questioned.

One significant alteration to the present system that is being used to some extent is the pass-fail system. Instead of five possible grades or levels of achievement, only two are present in this system, pass or fail. Yet other possible systems are being implemented or experimented with, such as

increasing the number of grading categories and imposing uniformity in the percentages of each of the grades given by different graders in the same system.

The Aeronautical Engineering Department of the Naval Postgraduate School is currently testing a new system of grading. This system replaces the usual letter grades and their numerical equivalents with "operational" grades. These "Operational" grades represent decisions as to whether a student's performance in a course warrants credit toward a specified degree. This new system was proposed to correct the deficiency in the letter grade-numerical system of the sometime unclear relationship between individual course performances by a student and the ultimate decision regarding granting of a degree, as well as other deficiencies.

B. SUMMARY OF LITERATURE CONCERNING GRADING

There has been a considerable amount of literature devoted to grading system questions. However, most of this literature takes the form of expression of opinion, with apparently little objective study being done. Many relatively unimportant issues are considered at length while basic issues, such as how well grades serve their intended purpose, appear to be almost ignored. In many cases where the important issues are considered, it is in an unobjective manner with inferences made from questionable data.

Explanations of grading fundamentals and other aspects of grading can be found in many educational and psychological measurement textbooks. Examples of such textbooks are:

Psychological Measurement and Prediction, by Paul Horst (5), Assessment of Behavior, by John E. Horrocks (4), and Measuring Educational Achievement, by Robert L. Ebel (1). Numerous articles are available in various educational journals concerning proposed changes in the present grading system. A good summary of such literature and a representative listing of it is found in the American Association for Higher Education Research, Research Report Number 3, by Johnathan R. Warren, entitled "Current Grading Practices," (10). The bibliography of this paper also contains several references on this subject.

There are few mathematical treatments of grading questions. In a paper titled, "Relations Between Grade Point Averages and Collegiate Course Grade Distributions," by Bruce G. Rodgers and William A. Mehrens (8), a model is proposed for predicting grade point average distributions from given grade distributions, allowing educational institutions to assess the possible effects of grading policy changes before actually implementing them. Robert L. Ebel, in a paper titled, "The Relation of Scale Fineness to Grade Accuracy," (2), states that "regardless of the inaccuracy of the basis for grading, the finer the scale used for reporting grades, the more accurate the grade reports will be." Ebel's method of proof of this statement is to present a single numerical example; it is not proven in general. This particular question will be addressed later in this paper.

C. THE PROBLEM

As mentioned above, there is a large amount written on opinions concerning grading questions, yet little about the theoretical aspects of grades. Grading and computation of grade point averages are mathematical processes and investigation of the subject in a mathematical light seems highly desirable. It is the intent of the author to examine various grading system characteristics, primarily from a mathematical point of view, with the aim of gaining insight into grading systems and procedures. Mathematical models of grading systems should aid in answering such important questions as, "How well do grading systems do what they should do?" and, "How will changes in grading system structure effect the effectiveness of the system?"

In this paper the variance of grade point average (GPA) is used as the measure of effectiveness in the basic model and its variations. The author feels this to be justified since GPA is often used in practice to differentiate between students and to rank them. The smaller the variance of the GPA of individual students, the more accurate the ranking according to GPA and the more informative the GPA, assuming unbiasedness in the statistical sense. Variance of GPA was chosen as the measure of effectiveness also because it could easily be modeled. Using this measure of effectiveness, various aspects of grading systems are modeled and the effects of certain parameter changes are observed. Rather than directly determine the benefits of particular grading systems,

in this paper the relationships between various parameters in the GPA process will be examined.

II. THE EFFECT OF VARIATION IN STUDENT PERFORMANCE LEVEL ON VARIANCE OF GPA

The purpose of the models used in this and following sections is to gain insight into the GPA process. Thus in some cases the models may be unrealistic and overlook true complications. This paper is also limited to consideration of the traditional letter grade-numerical equivalent GPA system. Despite these limitations, insight into various grading system questions can be obtained from the models.

In this section, the relationship between the "true" level of student performance and the variance of the GPA is modeled. To look at this relationship, the meaning of "true" level of student performance must be introduced.

A. TRUE STUDENT PERFORMANCE DISTRIBUTION

In a given course a student performs at some true level in relation to all others taking the same course. Suppose the grader determines what factors are to be considered in judging this performance and how they are to be combined into an overall performance measure. Although the grader may not be able to accurately determine the level of performance of the student on these factors, the student actually does perform at some level in each area considered. Combining the individual factor performances, a true level of overall course performance by the student is obtained.

Although the actual grading process may not occur exactly as stated above, it may be approximately, and it is

assumed in this model. Now consider a population of students that have been so graded. Each of these students has some true level of performance and thus (in theory) can be ranked according to their individual performance levels with respect to the particular grader's criteria. The student's performance places him at some position relative to all others who are in the population being considered and a centile point is thus determined.

Each student in the population is graded for a number of courses and has true performances which determine centile points for each course. Thus each student in the population has some true performance distribution over courses, measured in centiles, with a corresponding mean and standard deviation. Now it is desired to see how these parameters of the true student performance distribution effect variance of the GPA.

B. THE MODEL

To study the relationship between the true student performance distribution and the GPA, the following model is proposed.

Let T be a random variable which represents the true centile performance of a student in a given course. Before a grade is assigned to a student, his performance must be judged by a grader. The centile standing attributed to the student by the grader may differ from the student's true centile standing. Let E be a random variable representing the difference in true and grader determined centile standing or, in other words, the error introduced by the grader.

Let P represent the grader determined centile standing. From the discussion above it follows that this grader determined centile standing is the true centile standing combined with the grader error. Symbolically $P = T + E$, where E can take positive or negative values corresponding to positive or negative error, i.e., high or low judgment by the grader of true centile standing.

Assume that the distribution of grading errors, E, is such that the expected value of E is 0. The variance of E would depend on the particular distribution of E. For a given student the true performance level random variable, T, would also have an expected value, $E(T)$, and a variance $V(T)$. $E(T)$ is assumed to be some constant centile point, T_S, for a given student. This is the mean value of individual course true centile standings. $V(T)$ is a constant which depends on the distribution of T.

The expected value of P can be found as,

$$E(P) = E(T) + E(E) = T_S \; .$$

Assuming that T and E are independent random variables, the variance of P is found by the formula,

$$V(P) = V(T+E) = V(T) + V(E) \; .$$

The assumption that T and E are independent seems justified since it would be reasonable to assume that the distribution of errors in determination of centile performance by a grader is much the same all along the true centile performance range, except at the endpoints. Thus the random variable P, the grader observed performance, has a mean equal to

the expected true performance and has a variance which is the sum of the variances of the true performance and error distribution.

In the GPA process, the grader observed performance, P, for a given course is assigned some grade according to a grading mapping. Each of the grades has a particular numerical value. For example, one possible set of grades and their corresponding numerical values might be: A = 4.0, B = 3.0, C = 2.0, and D = 1.0. A student takes a number of courses, N, and the GPA is then computed by summing the product of the numerical values and course weights for each course and then dividing by the sum of the course weights. Course weights correspond to credit hours in the traditional system. This process can be expressed mathematically as,

$$GPA = \frac{\sum_{i=1}^{N} W_i \; M(P_i)}{\sum_{i=1}^{N} W_i}$$

where M is the grading mapping that assigns a particular grade and value to each possible outcome of P_i, the observed performance for the ith course, and W_i is the course weight of the ith course. An example of a possible mapping, M, might might be,

$$M(P_i) \quad \begin{array}{ll} A(4.0) & \text{if } 75 < P_i \leq 100 \\ B(3.0) & \text{if } 40 < P_i \leq 75 \\ C(2.0) & \text{if } 15 < P_i \leq 40 \\ D(1.0) & \text{if } \quad\;\; P_i \leq 15 \end{array}$$

Variation in course weights unnecessarily complicate manipulations of the model and do not add insight to the GPA process. Thus all courses will be considered of equal weight and the formula for determination of GPA simplifies to,

$$GPA = \frac{1}{N} \sum_{i=1}^{N} M(P_i) .$$

The expected value of the GPA, E(GPA), is easily found. From above,

$$E(GPA) = E[\frac{1}{N} \sum_{i=1}^{N} M(P_i)] = \frac{1}{N} E[\sum_{i=1}^{N} M(P_i)].$$

Assuming the $M(P_i)$'s to be identically distributed random variables,

$$E(GPA) = \frac{1}{N} \cdot N \cdot E[M(P_i)] = E[M(P_i)].$$

thus the expected GPA of a student is just the expected value of the grading mapping of the random grader observed performances.

For the variance of GPA the following is found,

$$V(GPA) = V[\frac{1}{N} \sum_{i=1}^{N} M(P_i)] = \frac{1}{N^2} V[\sum_{i=1}^{N} M(P_i)].$$

If the $M(P_i)$'s are assumed independent and identically distributed,

$$V(GPA) = \frac{1}{N^2} \sum_{i=1}^{N} V[M(P_i)] = \frac{1}{N^2} \cdot N \cdot V[M(P_i)]$$

$$V(GPA) = \frac{V[M(P_i)]}{N}$$

The $M(P_i)$'s may not be independent identically distributed random variables; however, in some systems it would be reasonable to assume that they are approximately so distributed. To make computations more tractable, the assumption of independence and identical distribution is made. Under these assumptions, the variance of the GPA is the variance of the grading mapping of the random grader observed performance divided by the number of courses.

Since E(GPA) and V(GPA) are functions of $E[M(P_i)]$ and $V[M(P_i)]$ respectively, it is necessary to examine these grading mapping parameters. Since $M(P_i)$ is a discrete random variable,

$$E[M(P_i)] = \sum_{k=1}^{m} p_k \, V_k \, ,$$

where m is the number of different possible grades in the mapping, p_k is the probability of receiving grade k given the distribution of P and V_k is the numerical value of the kth grade. Since $V[M(P_i)] = E[M(P_i)^2] = E[M(P_i)]^2$,

$$V[M(P_i)] = \sum_{k=1}^{m} p_k \, V_k^{\,2} - \sum_{k=1}^{m} \sum_{h=1}^{m} p_k \, V_k \, p_h \, V_h \, .$$

Now it can be observed that E(GPA) and V(GPA) are functions of the probabilities of obtaining each of the m grades in the grading mapping. Thus,

$$E(GPA) = \sum_{k=1}^{m} p_k \, V_k \qquad (1)$$

$$V(GPA) = \frac{\sum_{k=1}^{m} p_k V_k^{\,2} - \sum_{k=1}^{m} \sum_{h=1}^{m} p_k V_k p_h V_h}{N} \qquad (2)$$

C. THE RELATIONSHIP BETWEEN THE STUDENT PERFORMANCE DISTRIBUTION AND V(GPA)

For a fixed number of courses, N, and a given grading mapping, M, which determines fixed values for m, the number of different grades, and V_k, the numerical value of the kth grade, the p_k's, probabilities of receiving each of the grades, may be varied to effect V(GPA). Thus it is of interest to examine the relationship between the p_k's and the student performance distribution.

First the effect of varying E(P) is examined. Since it is assumed that E(E) = 0, then $E(P) = E(T) = T_S$ and examination of the effect of varying E(P) is equivalent to examination of the effect of varying E(T) or T_S. To isolate the effect of varying E(P) in a given grading mapping, M, a constant V(P) is assumed. The individual p_k's are functions of the distribution of P since they are the probabilities that P is in some specified centile range. Varying E(P) changes the distribution of P and thus the p_k's. An example is given to clarify this relationship.

Assume that for a given student P is uniformly distributed over the centile range [70,100]. Thus E(P) = 85 and V(P) = 75. With these results and the grading mapping used as an example in section B for the given grading mapping, the p_k's are obtained as follows; the probability that a given student receives the grade A in any one course, p_A, is given by,

$$p_A = p(75<P\leq 100) = \frac{25}{30} = .833.$$

Similarly for the other p_k's,

$$p_B = .167, \; p_C = 0 \text{ and } p_D = 0 \; .$$

Substituting these values into Equation 2 for V(GPA),

$$V(GPA) = \frac{.139}{N} \; .$$

Now let P be uniformly distributed over the interval [0,30] with E(P) = 15 and V(P) = 75. Using the same grading mapping the p_k's are now,

$$p_A = p_B = 0, \; p_C = .5, \text{ and } p_D = .5.$$

V(GPA) in this case is found to be,

$$V(GPA) = \frac{.29}{N} \; .$$

Thus it is observed that the V(GPA) can be different for different mean performances under a given grading mapping, with a constant student performance variance. Except for the case where only one possible grade could be received for any observed value of P, with consequent V(GPA) = 0, one would not expect a grading mapping to be of such a special form that V(GPA) would be constant over different mean performances. That is, it would seem unusual for the p_k's to change as mean performance changed in such a manner as to keep V(GPA) constant. Thus in general it would seem reasonable to say that variance of GPA is not constant over the range of mean student performance.

Now consider the variance of the observed student performance, V(P). To observe the effect of V(P) on E(GPA) and V(GPA), E(P) will be held constant in the example. Consider

the centile interval (35,65). Assume P is uniformly distributed over the interval, so E(P) = 50 and V(P) = 75. Using the grading mapping used in the previous example, the p_k's are found to be,

$$p_A = p_D = 0,\ p_B = .833,\ \text{and}\ p_C = .167.$$

From Equations 1 and 2 in Section B,

$$E(GPA) = 2.833 \text{ and } V(GPA) = \frac{.139}{N} .$$

Again let P be uniformly distributed over the interval [25,75]. Thus E(P) = 50 and V(P) = 208.33. In this case,

$$E(GPA) = 2.70 \text{ and } V(GPA) = \frac{.21}{N} .$$

It can now be observed that a different V(P) with the same E(P) and same grading mapping can yield different E(GPA) and V(GPA). An interesting conclusion to be drawn from the above is that a student's E(GPA) may be affected by the grader error distribution since V(P) = V(E)+ V(T). Taking courses under a grader with a given V(E), a student might have a different E(GPA) than if he had courses from a grader with a different V(E).

The above conclusions should hold in general since no unreasonable assumptions have been made and only in the cases where the distributions of the several variables and grading mapping have special characteristics would these conclusions not hold. The choice of grading mapping does not affect these conclusions since any other reasonable choice should give similar results. The assumption of a uniform distribution for P was only made to ease calculations

and does not affect the concept, any distribution should give similar results.

As is the case in the example, with a reasonable grading mapping V(GPA) would be expected to increase with increasing V(P) for midrange centile values of P. This increase is logical since increasing V(P) means that the distribution of P is getting less centralized which makes V(GPA) increase for reasonable grading mappings. Special problems, however, occur at the extremes of the range.

In reality E(GPA) and V(GPA) are complex functions of the T and E distributions and the grading mapping. To further explicitly define this relationship would require considerable assumptions concerning the distributions of T and E and about the grading mapping.

III. EFFECT OF MULTIPLE GRADERS

In the preceding section only one grading mapping (or, in effect, one grader) was assumed in determining individual course grade assignments. In general this is, of course, not the case and a number of graders apply different grading mappings to the centile performances of a student. In this section the effect of multiple graders on E(GPA) and V(GPA) will be examined.

A. THE MODEL

The multiple grader model presented in this section is derived from the single grader model in the previous section. With only one grader in the system, the formula for GPA was,

$$\text{GPA} = \frac{1}{N} \sum_{i=1}^{N} M(P_i) \ ,$$

where the variables are as defined in the previous section. Now for the multiple grader system, GPA is computed according to the formula,

$$\text{GPA} = \frac{1}{N} \sum_{k=1}^{L} \sum_{i=1}^{N_k} M_k(P_i) \ ,$$

where L is the number of graders in the system, N_k is the number of courses graded by the kth grader, $M_k(P_i)$ is the kth grader's mapping function, and N is the total number of courses taken.

B. RESULTS

For E(GPA) the following is found,

$$E(GPA) = E\left[\frac{1}{N} \sum_{k=1}^{L} \sum_{i=1}^{N_k} M_k(P_i)\right] = \frac{1}{N} \sum_{k=1}^{L} \sum_{i=1}^{N_k} E[M_k(P_i)] ,$$

using the properties of expected values of random variables. Again assuming that the course performances graded by a single instructor are identically distributed random variables then,

$$E(GPA) = \frac{1}{N} \sum_{k=1}^{L} N_k \; E[M_k(P_i)] .$$

When the number of courses graded by each grader is the same the expression reduces to,

$$E(GPA) = \frac{N_k}{N} \sum_{k=1}^{L} E[M_k(P_i)] = \frac{1}{L} \sum_{k=1}^{L} E[M_k(P_i)] .$$

This case is interesting since this could model the system where a student takes only one course from any single grader.

As would be expected, the E(GPA) in this multiple grader model is a linear combination of the expected values of the individual grader mappings for a given student. Thus a student's E(GPA) is just a weighted average of the E(GPA)'s for the student over the different graders. The student's E(GPA) could be different for the same true performance distribution, depending on the number of courses taken from each grader.

Now for the GPA in the multiple grader case,

$$V(GPA) = V[\frac{1}{N} \sum_{k=1}^{L} \sum_{i=1}^{N_k} M_k(P_i)] = \frac{1}{N^2} V[\sum_{k=1}^{L} \sum_{i=1}^{N_k} M_k(P_i)]$$

Assuming that the $M_k(P_i)$'s are independent,

$$V(GPA) = \frac{1}{N^2} \sum_{k=1}^{L} \sum_{i=1}^{N_k} V[M_k(P_i)].$$

The above assumption implies that the different grader mappings are independent of each other. This is probably not true in most cases but is assumed here to ease computations and since similar results should be obtained if the independence assumption is relaxed. Since the P_i's are assumed to be identically distributed,

$$V(GPA) = \frac{1}{N^2} \sum_{k=1}^{L} N_k \; V[M_k(P_i)]$$

For the case where an equal number of courses are taken from each grader, the expression for V(GPA) is,

$$V(GPA) = \frac{1}{N^2} N_k \sum_{k=1}^{L} V[M(P_i)] = \frac{1}{NL} \sum_{k=1}^{L} V[M_k(P_i)]$$

The V(GPA) for this multiple grader model is also a linear combination of the individual grader mapping function variances divided by N. This is interesting as it implies V(GPA) is averaged in this multiple grader process and no additional variance factor is added due to multiple graders.

In this case the student's V(GPA) would be somewhere between his high and low individual grader V(GPA), the exact value depending on the number of courses taken from each grader.

IV. A GRADING FINENESS MODEL

A. INTRODUCTION

The model in this section is proposed to examine the effect of grading scale fineness on the mathematical GPA system. In a given grading mapping or system a certain number of letter grades, or equivalently, their numerical values, can be received by the student. The question of interest is whether more or fewer grade categories are in some sense desirable. This question may be reasonably examined on some non-mathematical basis such as student motivation and ease of usage. However, the basis for this examination is the effect of variation in grading fineness on variance of GPA, where fineness is a function of the number of grade categories in a given grading mapping.

B. THE MODEL

In each course the student is assumed to perform at some centile point in relation to all other students as explained earlier. Suppose that this centile performance has a numerical equivalent on the GPA scale, whatever this GPA scale is. For example, the 85th centile might be equivalent to a 3.6 on a 0.0 to 4.0 GPA scale. Problems might arise in the actual determination of this mapping but such a mapping is assumed to exist. This assumption is made for ease in modeling and does not affect the result obtained.

Consider the case where a student receives a number of courses from a single grader. The grade received in a particular course, G, is then given by,

$$G = T + E_G + E_c ,$$

where T is the true student performance in GPA scale units, E_G is the grader error in GPA scale units, and E_c is the grading category error in GPA scale units given by,

$$E_c = (T + E_G) - G .$$

In other words, the grader observes a performance on the GPA scale which determines a certain numerical grade equivalent, according to the particular grading mapping. The differences between this grade equivalent and the observed performance is this grading category error.

To clarify this above system an example is given. Suppose a student's observed performance was 3.2 where the grading mapping was such that any performance between 2.5 and 3.5 received the numerical equivalent 3.0. Thus the grading category error is,

$$E_c = (T + E_G) - G = 3.2 - 3.0 = .2$$

Assume that T and E_G are independent random variables and can be combined into the grader observed performance random variable, P, so that,

$$G = P + E_c .$$

As before, if the expression for GPA is,

$$GPA = \frac{1}{N} \sum_{i=1}^{N} G_i ,$$

then,

$$GPA = \frac{1}{N} \sum_{i=1}^{N} (P_i + E_{c_i}) .$$

The expected value of GPA becomes,

$$E(GPA) = E(P) + E(E_c) ,$$

since the P_i's as well as the E_{c_i}'s, are assumed to be identically distributed random variables. The variance of GPA is,

$$V(GPA) = V[\frac{1}{N} \sum_{i=1}^{N} G_i] = \frac{1}{N} V[G_i] ,$$

or equivalently,

$$V(GPA) = \frac{1}{N} V[P + E_c]$$

Now the effect of a different number of grading categories on the category error, E_c, will be examined. Assume that the distribution of P is uniform over a portion of the GPA range for a given student and that this portion of the GPA range is divided into N_1 equal grading categories of length L_1. These assumptions are made to facilitate numerical analysis and their effect on the conclusions will be discussed in the next section. Consider another grading mapping with N_2 equal grading categories of length L_2. Suppose N_2 is greater than N_1. Since the total length covered by both systems is the same, L_2 must be smaller than L_1 or in other words, the second system has finer grading divisions. Also assume that the numerical equivalent for a particular grade, G, is the value at the midpoint of the grade category.

Consider an individual grade category from each of the two systems. Since P is uniform over the grade category and G is a constant in a given grade category, the expected value of E_c is zero in both systems. For the variance of E_c, again since P is uniform and G is a constant,

$$V(E_c) = V[P - G] = V[P] = \frac{L}{12} ,$$

where L is the length of the interval or in this case the length of the grading category. Since L_2 is smaller than L_1,

$$V(E_c)N_1 > V(E_c)N_2$$

and thus variance of E_c increases as the number of grading categories decrease for each of the individual grade categories. Since each grade category is of equal length within a system and since each is equally as likely to occur, the conditional distribution of E_c given P falls in any particular category is equivalent to the unconditional distribution of E_c over the entire system. Thus the variance of E_c increases as the number of grading categories decrease for the entire system in this model.

C. DISCUSSION

In the above model, with the given assumptions, $V(E_c)$ increased as the number of grading categories decreased. This result was suggested by a similar result in a paper titled "The Relation of Scale Fineness to Grade Accuracy," by Robert L. Ebel. The validity of the assumptions is important in determining the general applicability of this

result. The assumption that the grade equivalent value was located at the center of the grading categories affects the moments of E_c but was made to ease modeling and similar results should be obtained with different location. The assumption of equal length grading categories seems to be reasonable; similar results should be obtained with unequal categories when expected values are taken. The assumption of uniform distribution of P might be reasonable in some cases, a normal distribution or possibly other distributions might be more accurate in other cases. However, the results should be similar for various distributions. Thus it appears that $V(E_c)$ varies inversely with the number of grading categories in general, regardless of the distribution of P or the precise length of each grading category. This seems intuitively justifiable since with more grading categories more of the P distribution mass gets closer to zero error points and the largest possible errors become smaller.

In general, $E(E_c)$ is probably not zero and is only zero when the grading categories and the distribution of P interact in a special way. Since $E(GPA) = E(P) + E(E_c)$, the expected value of the GPA obtained can be different for the same observed performance distribution when a different grading mapping is used. It would again seem reasonable that $E(E_c)$ would decrease as the number of grading categories increases since E_c for points of the P distribution tend to get smaller as the number of categories increases.

Thus E(GPA) should get closer to the true performance mean with more grading categories.

Since,

$$V(GPA) = \frac{1}{N} V(P + E_c) = \frac{1}{N}[V(P) + V(E_c)],$$

V(GPA) will increase with increasing $V(E_c)$ for a given P distribution. Thus V(GPA) appears to increase with a decreasing number of grading categories in the model. The effect of increasing the number of grading categories is to make GPA more precise in the sense of reducing variance. It is interesting to note that increasing the number of grading categories decreases V(GPA) even though the grader cannot determine true performance any more accurately.

However, other considerations limit the extent of a possible increase in the number of grading categories in a grading system. This is particularly true for a grading system whose basic purpose is to recognize and record significant differences in actual achievement. In such a system it is important that the category size remain large enough to insure that the observed performance results in the correct grade with high probability. Thus even though GPA preciseness can be improved by increasing the number of grading categories, only a limited number of categories should be used.

V. ACTUAL GRADE DATA ANALYSIS

In this section actual grade data is examined with the purpose of supporting ideas presented earlier in the paper, to test the validity of some assumptions made in the paper and to gain further insight into the mathematical aspects of the grading process. The grade data used was taken from the records of selected students that attended the United States Naval Postgraduate School during the period 1966-1972.

A. VARIANCE OF GRADES RECEIVED VS. GRADE POINT AVERAGE

In Section II it was observed from the model that V(GPA) could be dependent to some degree on the mean student performance. To see if any corresponding dependence in actual grade data was present, the following analysis is presented.

In the analysis the GPA of the grades received and variance of the grades received by a student are compared. It seems reasonable to assume that the actual GPA obtained by students in a given grading system should correspond within limitations to their mean student performances. The variance of grades received corresponds to V(GPA) as follows. The variance of the grades received or sample variance, S^2, can be expressed symbolically as,

$$S^2 = \frac{\sum_{k=1}^{m} N_k V_k^{\,2} - \frac{\left(\sum_{k=1}^{m} N_k V_k\right)^2}{N}}{N},$$

Where N is the total number of courses, V_k is the numerical grade equivalent for the kth grade and N_k is the number of

grades of type k that the student receives. It seems reasonable to assume that for large N, the ratio N_k/N should approximate the P_k, probability of receiving grade k, for the student in a given system. Using this approximation the above formula can be expressed as,

$$S^2 = \sum_{k=1}^{m} P_k V_k^{\,2} - \sum_{k=1}^{m} \sum_{h=1}^{m} P_k V_k \; P_n V_n \; ,$$

which is the formula for V(GPA) as derived in Section II when divided by N, the total number of grades received. Thus examining the relationship of GPA of grades received and variance of grades received relates directly to the model in Section II.

Three sections of Operations Research students' grades were used as sources of data. One section was chosen at random from all the sections graduating in each of three years; 1970, 1971, and 1972. Data was used from students that did not graduate that were in these sections to insure that all grades given would be present. GPA and variance of grades received are computed for all students in these sections for the grades they received in their first four quarters. The number of courses taken during their first four quarters averaged approximately 16 with little variance.

The estimator used for the GPA of grades received is $\overline{X}$, where,

$$\overline{X} = \frac{\sum_{i=1}^{m} a_i \; V_i}{\sum_{i=1}^{N} a_i} \; ,$$

where a_i is the number of credit hours of the particular course. The above estimator is unbiased. A reasonable estimator of the variance of grades received is $S^{2'}$ where,

$$S^{2'} = \frac{\sum_{i=1}^{N} a_i (V_i - \overline{X})^2}{\sum_{i=1}^{N} a_i}$$

This estimator, $S^{2'}$, is a biased estimator, but may be made unbiased by the factor k where,

$$k = \frac{1}{1 - \dfrac{\sum_{i=1}^{N} a_i^2}{(\sum_{i=1}^{N} a_i)^2}} .$$

Let S^2 equal k $S^{2'}$.

The values of $\overline{X}$ and S^2 are now computed for the grades received by each of these 64 students and are plotted in Figure 1. The plots of points from the individual sections were so similar that no differentiation by section is made in Figure 1.

From Figure 1 it is apparent that variance of the grades received is not independent of the GPA obtained in this grading system. This dependence is very pronounced at the high end of the $\overline{X}$ scale where the S^2's are all low in comparison with those for other ranges of $\overline{X}$.

Some correlation between $\overline{X}$ and S^2 is expected since very low S^2's occur only when the student receives predominantly one grade which could only occur in a certain small $\overline{X}$ range. Similarly, relatively high S^2's will occur only in $\overline{X}$ ranges

which result from combinations of grades equivalents of widely separated grades in the system. This effect is not as pronounced as might be expected due to the fact that most of the observed $\bar{X}$'s lie in a relatively small portion of the $\bar{X}$ range. This tends to reduce the obviousness of the dependence of S^2 on $\bar{X}$, but it is apparent, particularly in the high $\bar{X}$ range.

Thus from the data it seems reasonable to conclude that V(GPA) for a student depends to some degree on the student's mean performance level as was concluded in Section II. The degree of this dependence would depend on the particular grading system and the distribution of grades in the system. It should also be remembered that V(GPA) varies inversely with the number of courses taken and consequently the difference between S^2's due to this dependence on $\bar{X}$ would decrease with increasing N.

It is interesting to observe that actual limits exist for values that S^2 may attain over the $\bar{X}$ range. The factor k will be assumed to be unity which corresponds to using $S^{2'}$ as the estimator. These limits are indicated by dotted lines on Figures 1 and 2. As can be seen by comparing Figure 1 to Figure 2, the observed data points all lie in the low range of possible S^2's. This is reasonable since being in the high range implies a significant number of low grades present which is not the case.

The high limits are determined by plotting the $S^{2'}$'s for many different combinations of A's (4.0's) and X's

(0.0's) which are the $S^{2'}$ limiting grade combinations. With an infinite number of courses, any grade combination would result in a GPA that could be expressed equivalently as some combination of A's and X's. The A and X combination would always have the highest value for $S^{2'}$ since it would be the most dispersed combination possible. The lower limits are obtained by plotting various combinations of grades that are adjacent to each other in the system. Any other combination having the same $\bar{X}$ would have a higher $S^{2'}$ since it would result from grades that are more dispersed than the adjacent grade combination. The absolute minimum limits occur at the grade equivalent points themselves since the minimum variance is zero variance corresponding to all one grade.

B. VARIANCE OF GRADES RECEIVED VS. TIME

It seemed possible to the author that the variance of grades received might depend to some extent on the time that the student had been in the system. Any such dependence would directly affect the assumptions made earlier in the paper. To look for any possible dependence and to give general insight in this area, the following analysis is presented.

The unbiased variances of grades received, $S^{2'}$, by students in a single quarter are computed as in Section VA for the previously mentioned sections of students as well as for a fourth section of students. This fourth section is composed of 27 former Chemistry Department students at the

Postgraduate School from 1966 to 1972. These 27 students represent all the students that majored in Chemistry during this period and whose records were available.

The resulting data was pooled by section for each quarter and the estimator for the population variance, $\overline{S}^2$ was obtained by averaging the individual unbiased estimators. $\overline{S}^2$ is plotted in Figure 3 for the four sections for the students' first 8 quarters of study. $\overline{S}^2$ was also calculated for each quarter for all sections combined and is plotted on Figure 3.

From the plot on Figure 3 it can be seen that no apparent trend in $\overline{S}^2$ vs. time is present. Significant differences occur for different sections in different quarters but no overall trend is present. These individual quarter observed variance differences can be partially explained by the presence of significantly different course grade distributions in different quarters. This effect is present because the students in each section took very similar subjects, and in many cases exactly the same ones, in each quarter and thus a difference in observed $\overline{S}^2$ would be expected. The Chemistry Department section seems to have larger overall variance. This is partially due to the fact that the department's grade distribution was more varied than was that of the Operations Research Department.

In summary, no apparent time dependence is present. Thus the assumptions made earlier still seem tenable, with respect to time dependence.

C. VARIANCE OF GRADES RECEIVED WITH RESPECT TO DEPARTMENT SIZE

It seemed reasonable to the author that students from a small department with a small number of graders might obtain smaller observed variances in grades received than students from a larger department. This might occur because the student from the small department might be relatively more well known by the graders than a student in a larger department and some grader bias could develop with respect to the student. It seems unlikely for a significant amount of grader bias to develop for students in a large department. Other system peculiarities might be present which could cause a significant difference in variance of grades received for students in different departments. If there were such a difference present for any reason, it would imply that some type of dependence was present. If so the assumptions made earlier in this paper would be adversely affected. For this reason the following data analysis is presented.

Data points were obtained for the large department from students in the three previously mentioned Operations Research Department sections and for the small department from students in the previously mentioned Chemistry Department section. Although no actual figures are presented, the Chemistry Department students did have fewer graders than the Operations Research Department students, on the average, while taking approximately the same number of courses.

The unbiased estimator, S^2, is computed as before for the grades received by each of the students in their 5th through 8th quarters. If this bias is present, it should show up primarily in the latter quarters and thus 5th through 8th quarter grades are the ones used. The obtained S^2's are plotted vs. the observed GPA in Figure 4.

To test for a significant difference between departments the analysis of covariance technique was used. This method was used to remove the effect that relatively different numbers of points from different $\overline{X}$ regions might have since the observed S^2's are dependent on $\overline{X}$.

However, neither of the F ratios obtained in the analysis of covariance, (corrected for differences in $\overline{X}$ and uncorrected for differences in $\overline{X}$) were statistically significant. Thus the hypothesis of equal means of observed variances in grades received could not be rejected.

This conclusion can be verified by examining the plotted points in Figure 4. It seems apparent from this plot that both sets of data are from the same population. Thus no evidence of dependence due to department size or grader bias is found. The independence assumptions made earlier in this paper seem not unreasonable with respect to grader bias and department size.

VI. SUMMARY AND CONCLUSIONS

In this section the major conclusions obtained will be stated, significant problems encountered by the author will be mentioned, and suggestions for further study in the grading system modeling area will be proposed.

A. CONCLUSIONS

Many questions have arisen concerning grading systems. From a review of the current literature it is apparent that very little mathematical analysis has been completed in this area. Since the grading process as most often used is a mathematical process, it seems desirable to examine grading systems from the mathematical point of view.

From the model presented in Section II it is observed that the expected value of the GPA obtained by a student can vary with a changing value of performance distribution variance even though the expected value of the performance distribution remains constant. The variance of the GPA obtained not only changes due to performance distribution variance difference for a constant mean performance level but also can change with different mean performance levels for a constant performance distribution variance. Thus a difference in the "precision" of the GPA for different students seems likely to exist.

From the model presented in Section III it is shown that the E(GPA) and V(GPA) received by students from multiple

graders are linear combinations of the E(GPA) and V(GPA) of the individual graders respectively. Thus the limits of these multiple grader parameters are determined by the maximum and minimum values from the individual graders.

In Section IV it is shown that V(GPA) can be reduced by increasing the number of grading categories no matter how inaccurate the grade determination by the grader may be. Increasing the number of grading categories thus seems desirable but must be implemented cautiously as other factors may be adversely affected by this increase.

The data analysis in Section V confirmed that V(GPA) is not independent of mean student performance and the GPA received. No significant dependence of V(GPA) to time in the grading system or department size were found. Thus the independence assumptions made in earlier sections actually seemed consistent with the data findings.

B. PROBLEMS ENCOUNTERED

As is the case in any mathematical modeling, many problems were encountered during the preparation of this paper. Following below are some of the major problems.

One major problem is to accurately define and model the student performance distribution. This is a necessary step to accomplish since this distribution is the "input" to the model and is what is hopefully being measured. To see how accurately it is being measured it must first be described. This is not easily accomplished since this distribution can only be observed indirectly.

Similarly it is a problem to model and evaluate the parameters of the grader error distribution. They too can only be observed indirectly. However, to separate error due to graders and that due to the grading system some assumptions must be made about this distribution.

The dependence of course grades is another not easily reckoned with problem. It seems very likely that many types of dependencies exist, yet to determine to what degree is not easily accomplished. It is very hard to infer reasons for correlations from data analysis and to model dependence when it does exist.

C. AREAS FOR FURTHER STUDY

Since little actual mathematical analysis has been done in this area, many avenues of pursuit are open. Presented below are several areas the author feels to be worthwhile study areas as well as somewhat manageable.

There is a need for additional actual grade data analysis. This analysis could be directed to many areas of the problem. Without much modeling it seems possible to gain a large amount of insight into grading systems through data analysis.

The question of whether or not an optional grading system exists is worthwhile. This analysis could also take the form of comparing alternative systems with some measure of effectiveness.

Looking at the grading system with respect to its ranking function also has possibilities. Comparison could be made

between ultimate GPA ranking and individual course rankings, combined into an overall ranking.

Extensions of and revisions to the several topics covered in this paper also seem worthwhile.

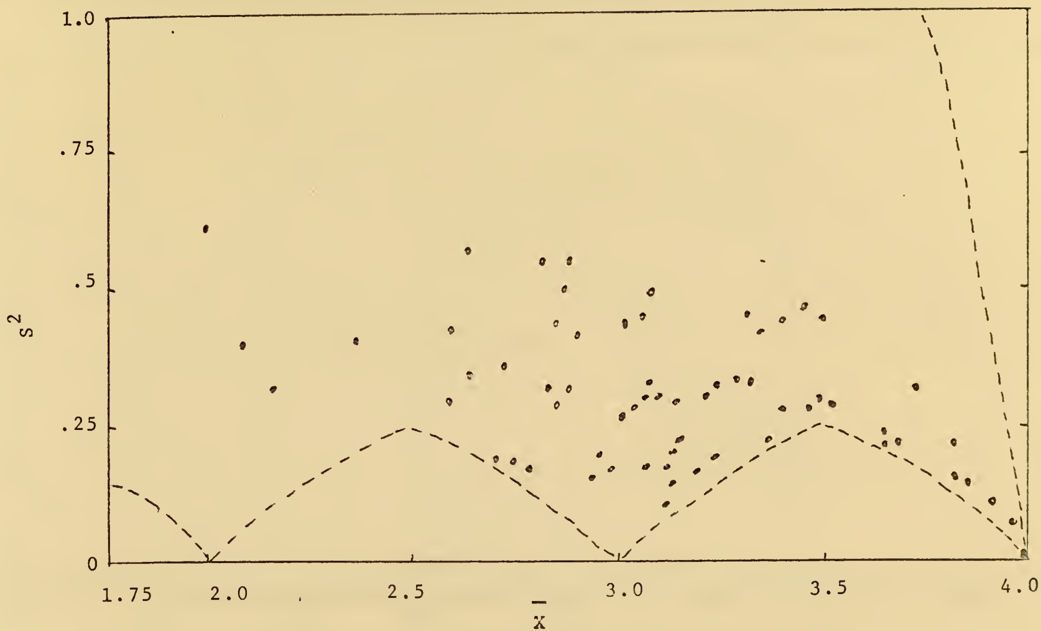

Figure 1. Observed grade unbiased variance vs. observed grade mean.

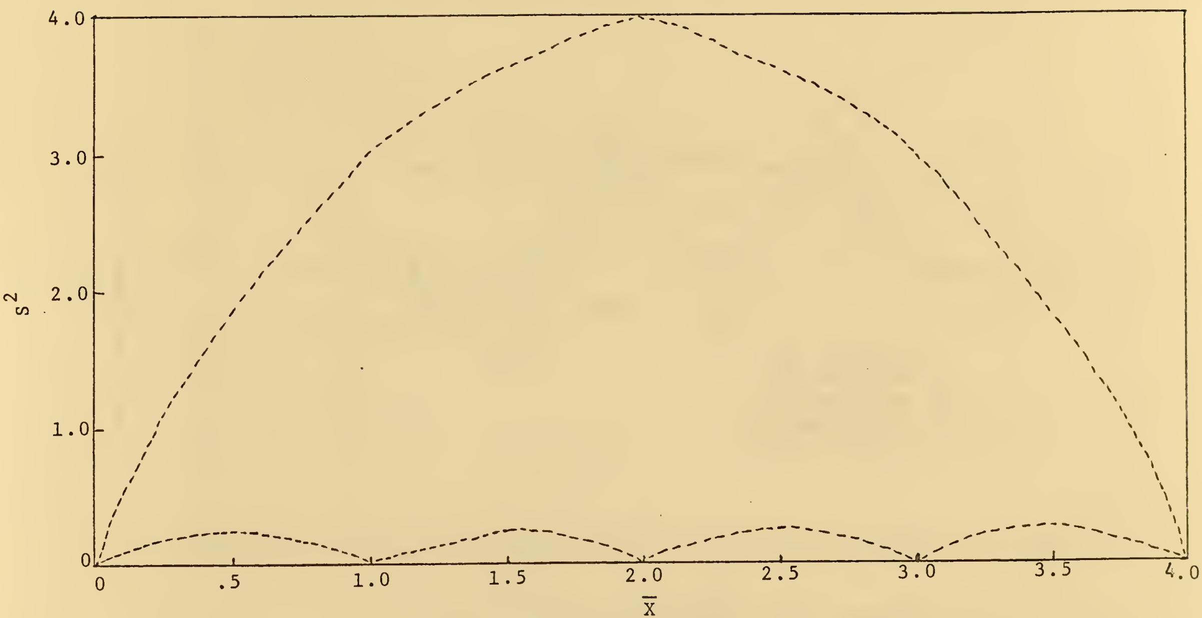

Figure 2. Maximum and minimum possible grade variance vs. observed grade mean.

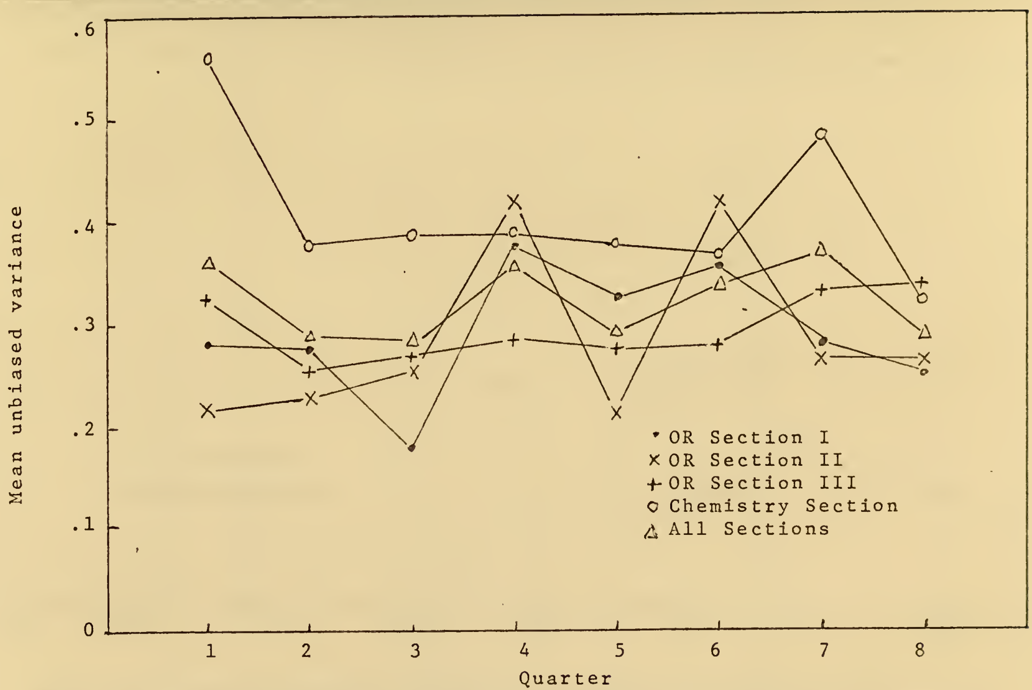

Figure 3. Mean unbiased grade variance vs. quarter.

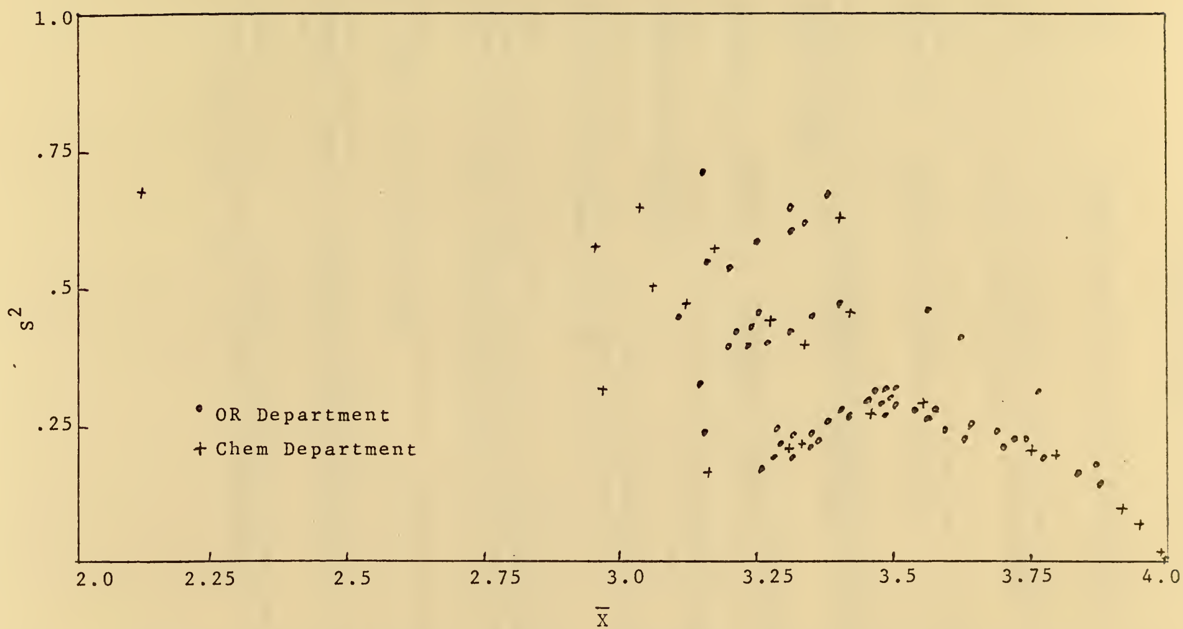

Figure 4. Observed grade unbiased variance vs. observed grade mean for students from two different departments.

BIBLIOGRAPHY

1. Ebel, R. L., Measuring Educational Achievement, Prentice-Hall, Inc., 1965.

2. Ebel, R. L., "The Relation of Scale Fineness to Grade Accuracy," Journal of Educational Measurement, v. 6, No. 4, p. 217-221, Winter 1969.

3. Elbow, P. H., "More Accurate Evaluation of Student Performance," Journal of Higher Education, v. 40, p. 219-230, March 1969.

4. Horrocks, J. E., Assessment of Behavior: The Methodology and Content of Psychological Measurement, C. E. Merrill Books, 1964.

5. Horst, P., Psychological Measurement and Prediction, Brooks-Cole Division, Wadsworth Publishing Co., 1966.

6. Hoyt, D. P., "Nationality and the Grading Process," Educational Record, v. 51, p. 305-309, Summer 1970.

7. Leisenring, L. B., and others, "Evaluating Evaluations; Grading Practices," College and University, v. 45, p. 349-359, Summer 1970.

8. Mehrens, W. A. and Rogers, B. E., "Relations Between Grade Point Averages and Collegiate Course Grade Distributions," Journal of Educational Research, v. 64, No. 4, p. 169-171, December 1970.

9. Simon, S. B., "Down with Grades," Todays Education, v. 58, p. 54, April 1969.

10. American Association for Higher Education Research Report Number 3, Current Grading Practices, by J. R. Warren, p. 3-6, 15 January 1971.

11. Wilson, K. M., "Increased Selectivity and Institutional Grading Standards," College and University, v. 46, p. 46-53, Fall 1970.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Professor Donald R. Barr Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	5
4.	Professor Theodore A. Gawain Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
5.	Ensign David Edward Polzien, USN 5960 W. Buzzell Road Gladwin, Michigan 48624	5

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

REPORT TITLE

Characteristics of Grading Models

DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Master's Thesis; December 1972

AUTHOR(S) (First name, middle initial, last name)

David E. Polzien

REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1972	50	11

a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

ABSTRACT

Many questions have arisen concerning grading systems. From a review of the current literature, it is apparent that very little mathematical analysis has been completed in this area. Since the grading process as most often used is a mathematical process, it seems desirable to examine grading systems from the mathematical point of view.

In this paper the familiar five-letter grading system is modeled mathematically. Variance of the grade point average is used as the measure of effectiveness. The grading system and its parameters are defined and the effect of variation in the parameters of the student performance distribution is observed. The effect of having multiple graders in a grading system is also modeled as is the effect of changing the number of grading categories in a grading system.

Grade data was obtained from the records of a number of students who attended the Naval Postgraduate School and an analysis of this data is presented.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
mathematical model						
grading systems						
grade point average						

Thesis 143216
P734 Polzien
c.1 Characteristics of
grading models.